

ABSTRACT

Disclosed are a variety of methods and computer systems for use in the analysis of gene and protein expression data. Also disclosed are methods for the definition of the cellular state of cells and tissues from multidimensional physiological data such as those obtained from gene expression measurements with DNA microarrays. A variety of classification methods can be applied to expression data to achieve this goal. Demonstrated is the application of several statistical tools including Wilks' lambda ratio of within-group to total variance, Fisher Discriminant Analysis, and the misclassification error rate to the identification of discriminating genes and the overall classification of expression data. Examples from several different cases demonstrate the ability of the method to produce well-separated groups in the projection space representing distinct physiological states. The method can be augmented and is useful in disease diagnosis, drug screening and bioprocessing applications.